# Systems Biology and its Role in Predictive Health and Personalized Medicine

**Eberhard O. Voit**

Integrative BioSystems Institute (IBSI)
W. H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology and Emory University
Atlanta, Georgia

Distinguished Lecture Series in Systems Biology, GT
February 5, 2008

# Overview

Summary in three easy slides (no math, no theory)

---

Limitations of traditional medicine

Requirements for predictive and personalized medicine

Potential of systems biology

Canonical modeling

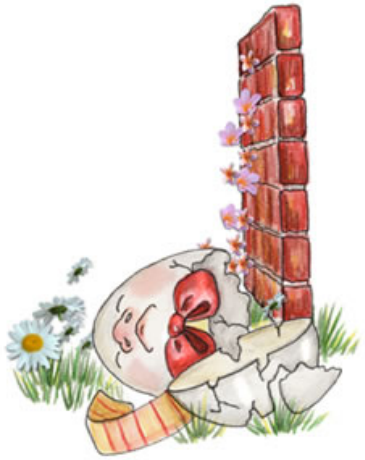
Illustration: Parkinson's disease (preliminary)

Challenges and state of the art

# Limitations of Traditional Medicine

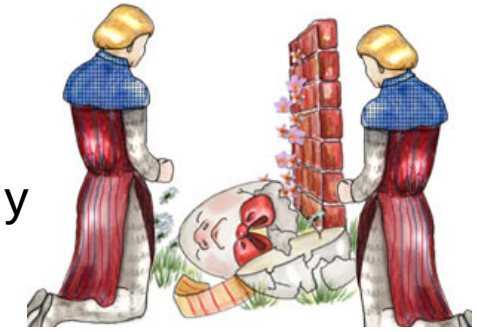
Humpty Dumpty
Sat on a wall.

Humpty Dumpty
Had a great fall.



All the King's horses
And all the King's men

Couldn't put Humpty
Together again.



# Limitations of Traditional Medicine??

Leslie Wesley
Lived a good life,

But started to tremble
At age 65.

All the cool gadgets
And medicine men

Couldn't make Leslie
Healthy again.

# Claim: ystems Biology May Help!

# How does Medicine Work?

***Injury (e.g., bone fracture):***

Fix problem approximately; nature takes care of rest

***Chronic Disease (e.g., diabetes, Parkinson's):***

- Symptoms evident; root causes often not
- Epidemiology associates disease with risk factors or biomarkers (high BP, cholesterol, …), based on statistics, *using population averages*
- Lab Science associates disease with mechanisms, based on animal experiments, *using population averages*

# What's the Problem with Averages?

Average input information allows

***average predictions*** at best

But: Our goal is to make predictions for

***specific individuals***

- with respect to personal health risks (predictive health)
- with respect to personal diagnostics and treatment (personalized medicine)

# What is Required Then?

***Need to progress from average input-output correlations to a deeper understanding of disease processes in individuals***

***Challenges:***

Get data

Analyze them appropriately

***Hope: Analogy with engineering***

We do not need to take apart every machine we encounter, if we understand the principles that make this type of machine functional.

# What is Required Then?

***Need to progress from average input-output correlations to a deeper understanding of disease processes in individuals***

***Challenges:***

Get data

Analyze them appropriately

***Hope: Analogy with engineering***

We do not need to take apart every machine we encounter, ***if we understand the principles*** that make this type of machine functional.

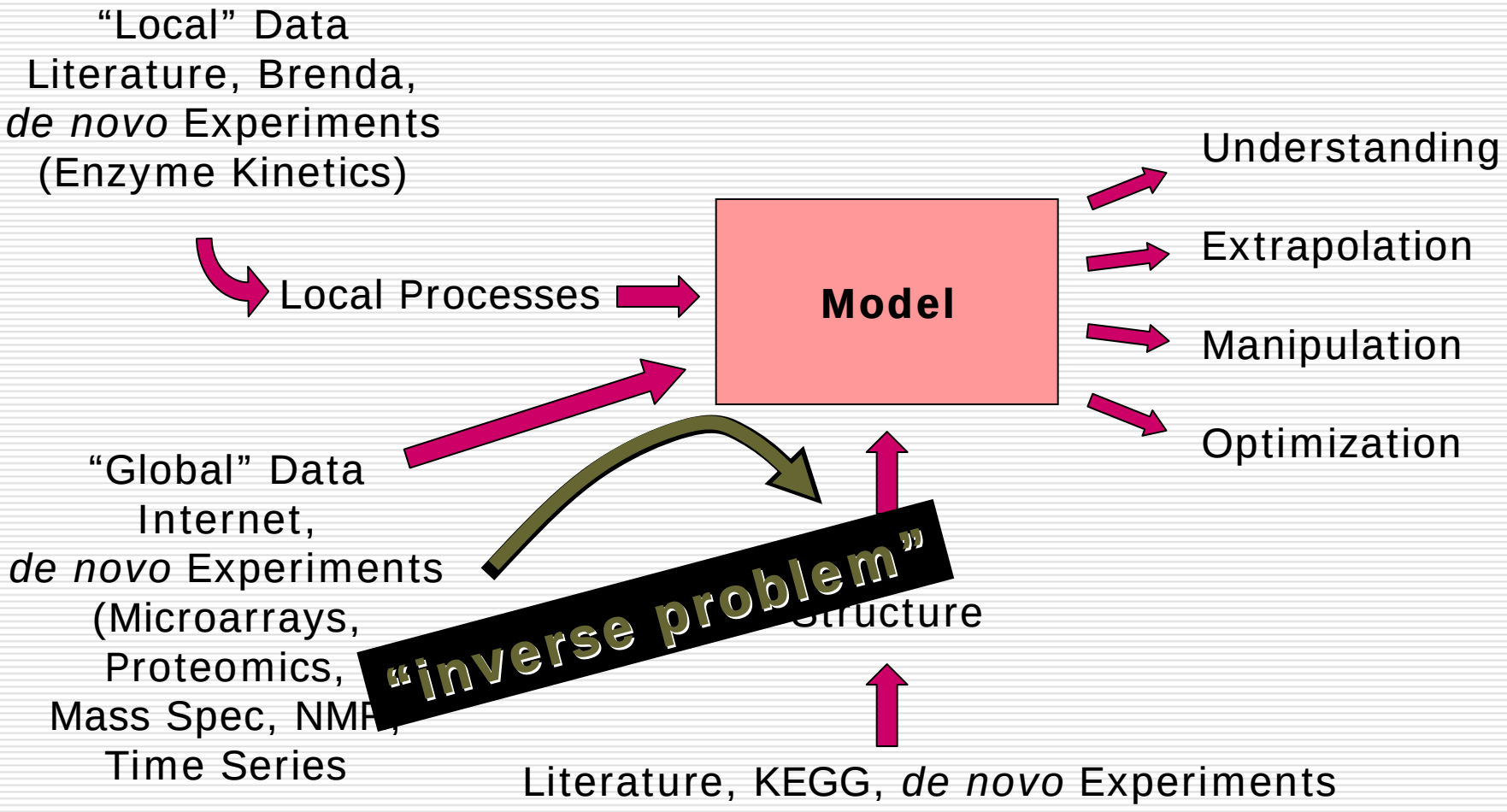
# Understanding Design Principles In Biology and Medicine

Instructions Manual for Systems Biology

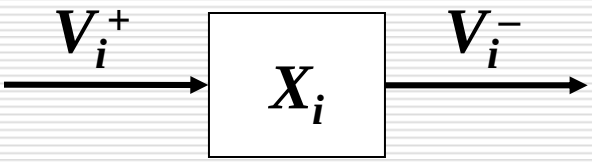
***Strategy:***

- "Under the hood"
- Collect, identify, characterize all components
- Quantify processes and interactions
- Put all into integrative models
- Study average behavior of these models
- Study effect of each component, process
- Trace abnormalities (diseases) to possible root causes
- Develop (in the model) strategies of counteraction
- Validate and use

# At the Center: a Model



# Formulation of a Model for Complex Systems




$$\dot{X}_i = \frac{dX_i}{dt} = V_i^+ - V_i^-$$

$$V_i^+ = V_i^+(\underbrace{X_1, X_2, ..., X_n}_{\textit{inside}}, \underbrace{X_{n+1}, ..., X_{n+m}}_{\textit{outside}})$$

**complex**

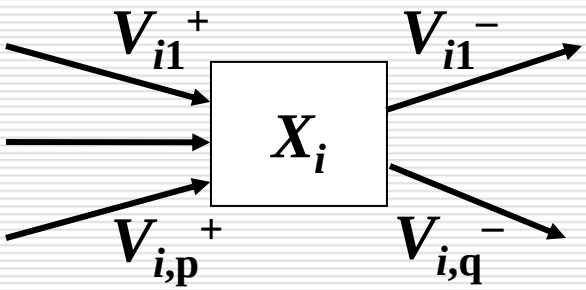
Michael Savageau, 1969: Thou shall approximate in log space!

Result: Power-law terms; Biochemical Systems Theory; "Canonical Modeling"

# Alternative Formulations Within BST

**S-system Form:**

$$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \ldots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \ldots X_{n+m}^{h_{i,n+m}}$$

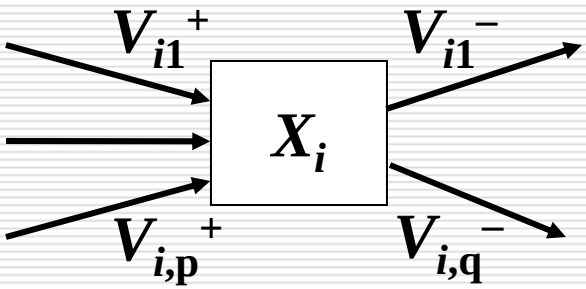



$$\dot{X}_i = \frac{dX_i}{dt} = \sum V_{ij}^+ - \sum V_{ij}^-$$

# Alternative Formulations

## S-system Form:

$$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \ldots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \ldots X_{n+m}^{h_{i,n+m}}$$



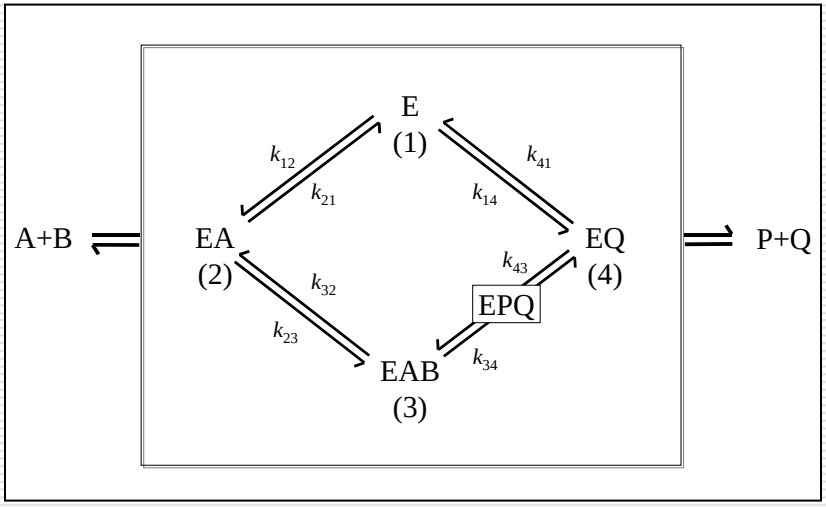

$$\dot{X}_i = \frac{dX_i}{dt} = \sum V_{ij}^+ - \sum V_{ij}^-$$

## Generalized Mass Action Form:

$$\dot{X}_i = \sum \pm \gamma_{ik} \prod X_j^{f_{ijk}}$$

# Why not Use "True" Rate Functions

$$A+B \rightleftharpoons P+Q$$




$$v = \frac{\left(\frac{\text{num.1}}{\text{coef. AB}}\right)(A)(B) - \left(\frac{\text{num.1}}{\text{coef. AB}} \times \frac{\text{num.2}}{\text{num.1}}\right)(P)(Q)}{\left(\frac{\text{constant}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right) + \left(\frac{\text{coef. A}}{\text{coef. AB}}\right)(A) + \left(\frac{\text{coef. B}}{\text{coef. AB}}\right)(B)}$$
$$+ \left(\frac{\text{coef. AB}}{\text{coef. AB}}\right)(A)(B) + \left(\frac{\text{coef.P}}{\text{coef. AP}} \times \frac{\text{coef. AP}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right)(P)$$
$$+ \left(\frac{\text{coef.Q}}{\text{constant}} \times \frac{\text{constant}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right)(Q)$$
$$+ \left(\frac{\text{coef.AP}}{\text{coef. A}} \times \frac{\text{coef.A}}{\text{coef. AB}}\right)(A)(P) + \left(\frac{\text{coef.BQ}}{\text{coef. B}} \times \frac{\text{coef.B}}{\text{coef. AB}}\right)(B)(Q)$$
$$+ \left(\frac{\text{coef.PQ}}{\text{coef.Q}} \times \frac{\text{coef.Q}}{\text{constant}} \times \frac{\text{constant}}{\text{coef. A}} \times \frac{\text{coef. A}}{\text{coef. AB}}\right)(P)(Q)$$
$$+ \left(\frac{\text{coef.ABP}}{\text{coef. AB}}\right)(A)(B)(P)$$
$$+ \left(\frac{\text{coef.BPQ}}{\text{coef. BQ}} \times \frac{\text{coef.BQ}}{\text{coef. B}} \times \frac{\text{coef.B}}{\text{coef. AB}}\right)(B)(P)(Q)$$

*from Schultz (1994)*

# Realistic Situation: Where to Start?

Task: Assess with a model the importance of genetic predisposition in Parkinson's Disease

A priori, no guidelines for model design

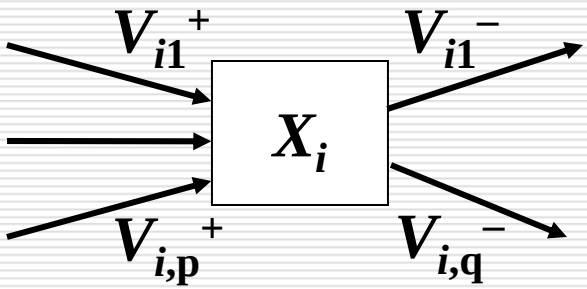
Physical (first) principles too far removed

Gene expression too coarse

Little overlap between metabolic and neurological modeling

How to incorporate "soft" clinical information:
"Patient has problems with balance"

# Canonical Models: No Panacea, But Predictable and Locally Guaranteed




$$\dot{X}_i = \frac{dX_i}{dt} = \sum V_{ij}^{+} - \sum V_{ij}^{-}$$

**S-system:** $$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \ldots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \ldots X_{n+m}^{h_{i,n+m}}$$

**GMA Form:** $$\dot{X}_i = \sum \pm \gamma_{ik} \prod X_j^{f_{ijk}}$$

Which variable affects which flux?
How strongly? Influence positive or negative?
Model guaranteed to be locally good.

# Advantage: Standardized Analyses

Model Design: Strict guidelines

Domain of Representation: All nonlinearities possible

Steady-State Analysis: Existence, stability, sensitivities

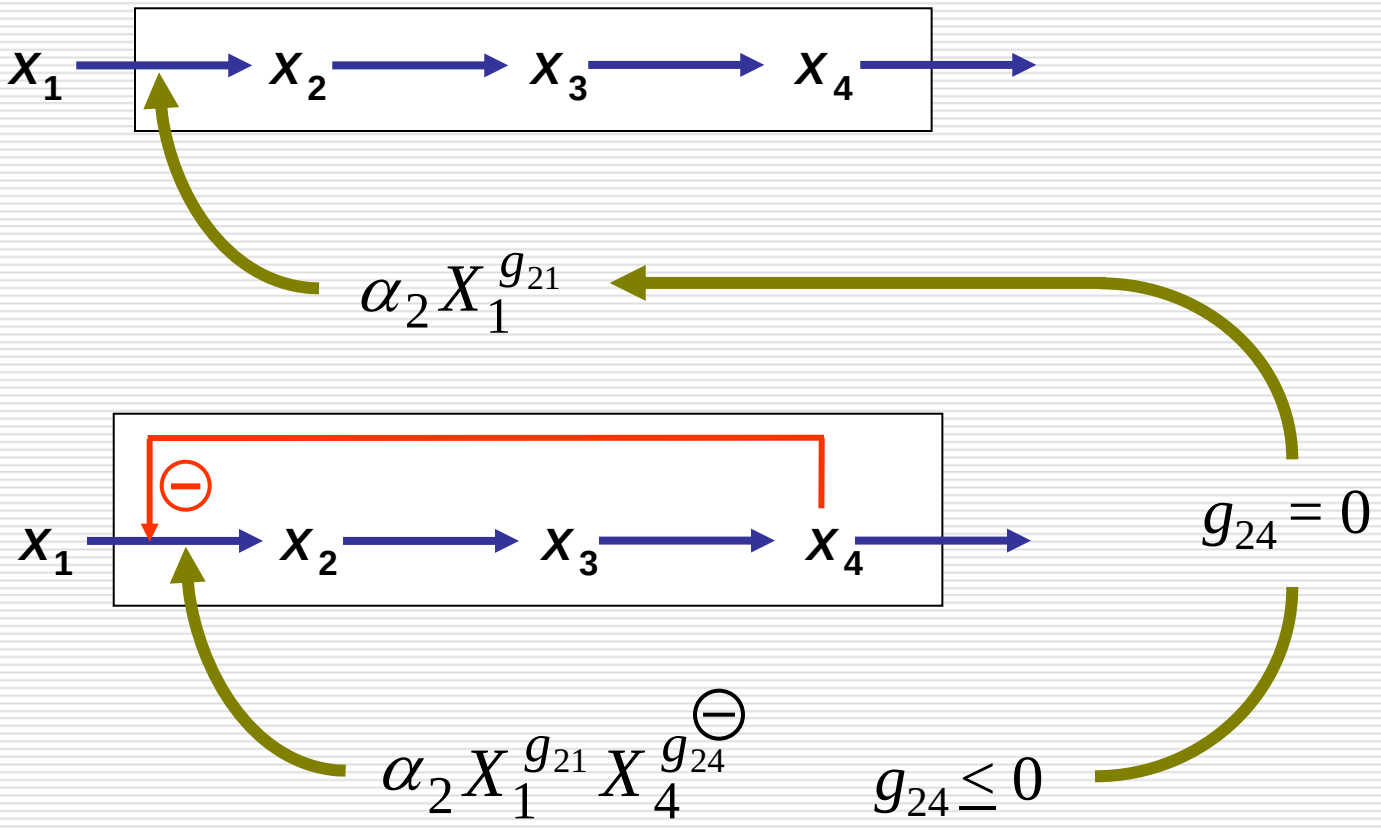
Dynamics: Optimized, customized software tools

Optimization (*e.g.*, Metabolic Engineering): Often linear

Design and Operating Principles: Powerful methods

Model Results: Clear interpretation

# Unique Meaning of Parameters



# Applications

Pathways: purines, glycolysis, citric acid, TCA, red blood cell, trehalose, sphingolipids, ...

Genes: circuitry, regulation,...

Genome: explain expression patterns upon stimulus

Growth, immunology, pharmaceutical science, forestry, ...

Metabolic engineering: optimize yield in microbial pathways

Dynamic labeling analyses possible

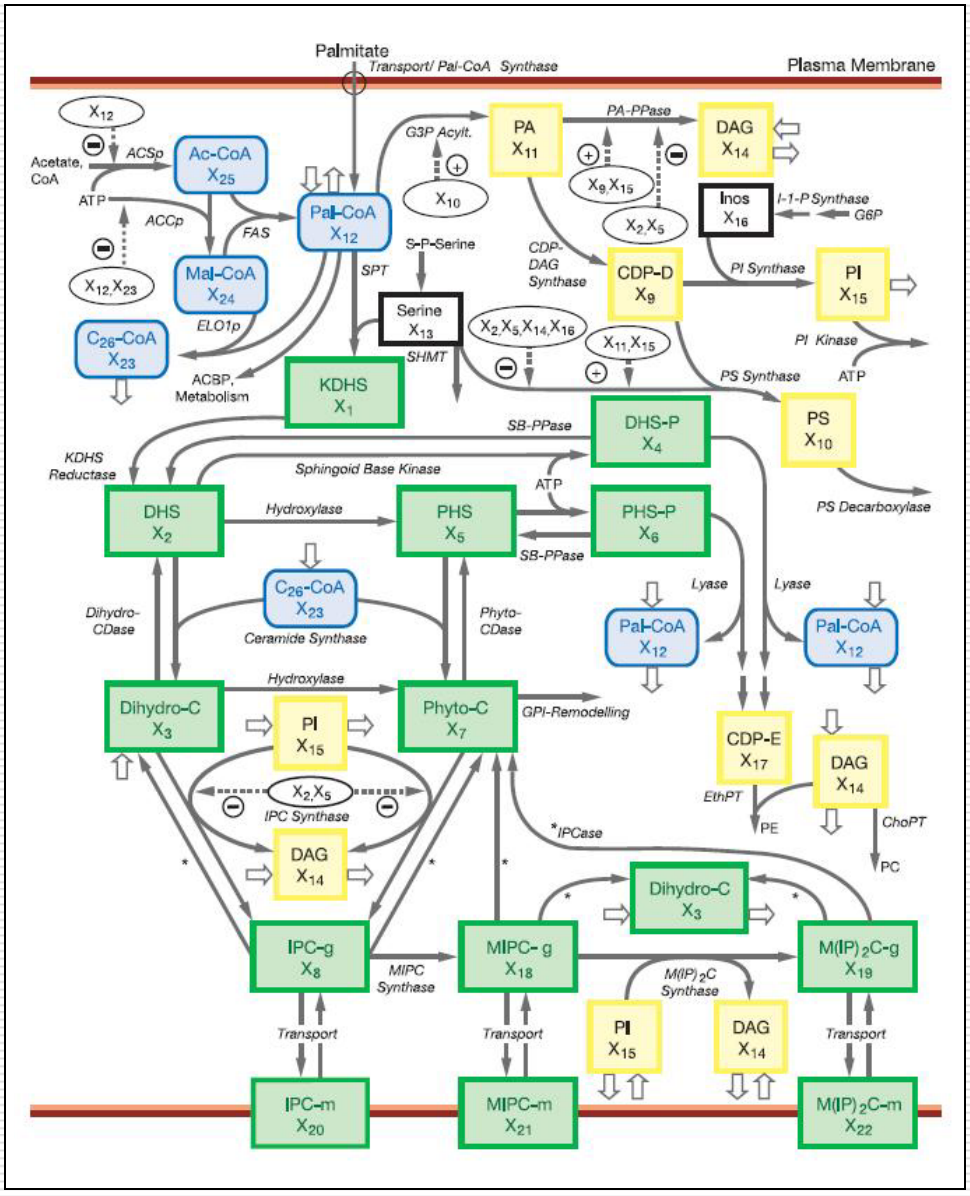
Math: recasting, function classification, bifurcation analysis,...

Statistics: S-system representation, S-distribution, trends; applied to seafood safety, marine mammals, health economics

# Doable Size

***Sphingolipid pathway (purely metabolic)***

1. Many metabolites
2. Many reactions
3. Many stimuli and agents regulate several enzymes of lipid metabolism
4. Some *in vivo* experiments




*Alvarez, Sims , Hannun, Voit*
*JTB*, 2004; *Nature*, 2005

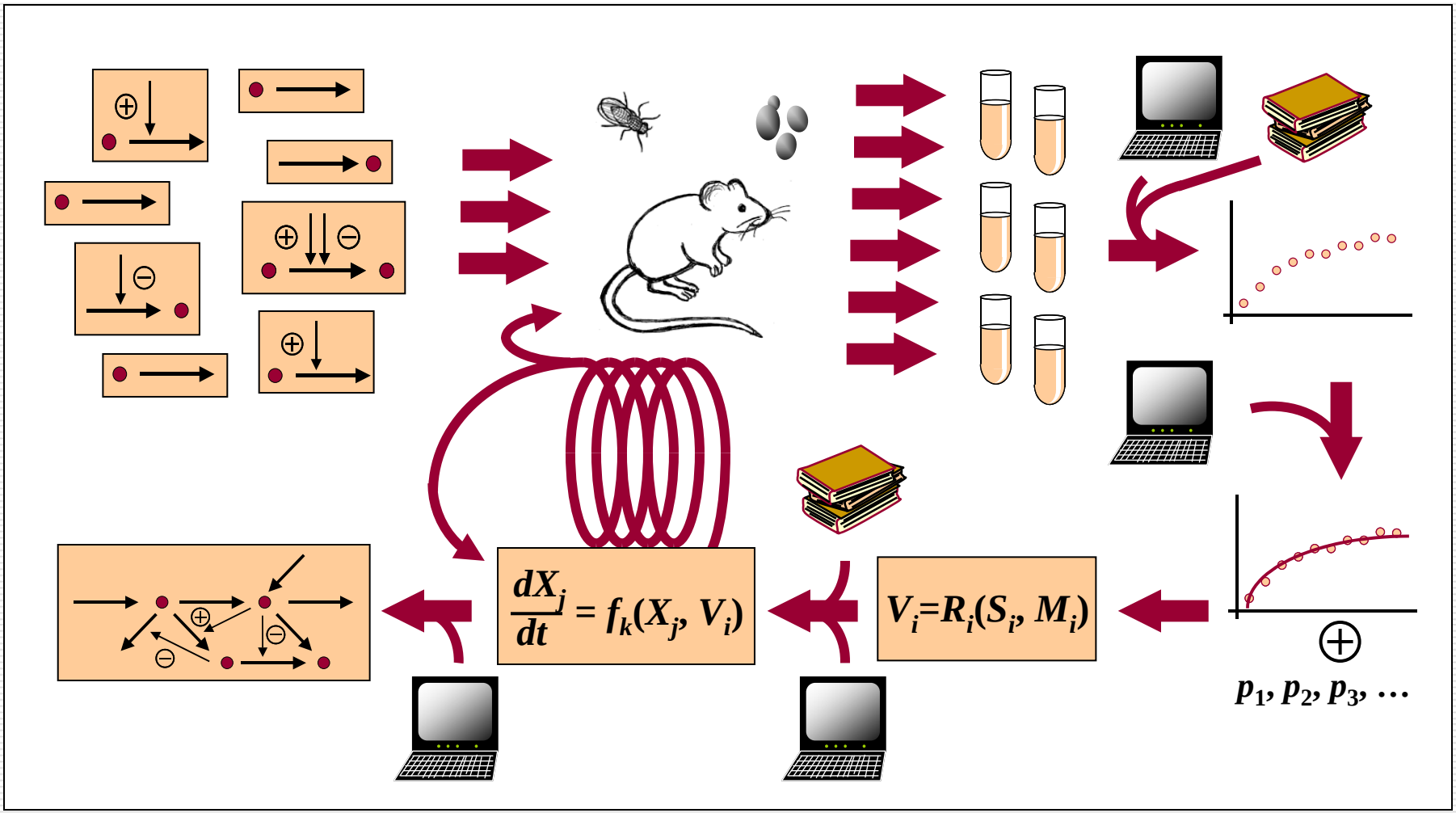
# Crucial Issue

Estimation of parameter values is complicated and time consuming

Typical procedure:

- Characterize every step (process) biologically
- Formulate as power-law function
- Determine parameter values
- Merge all process descriptions
- Check system description against “global” data

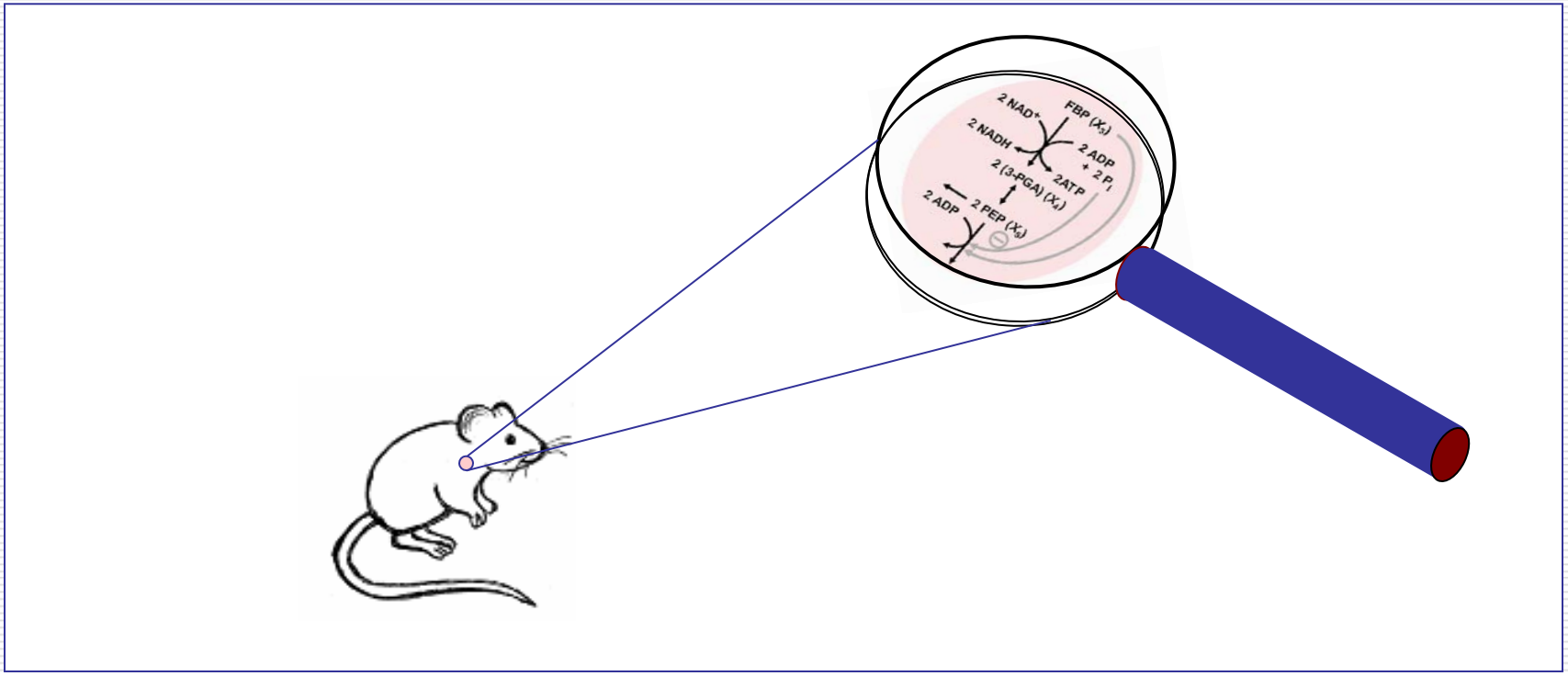
# Flow Chart of Traditional Modeling Strategy



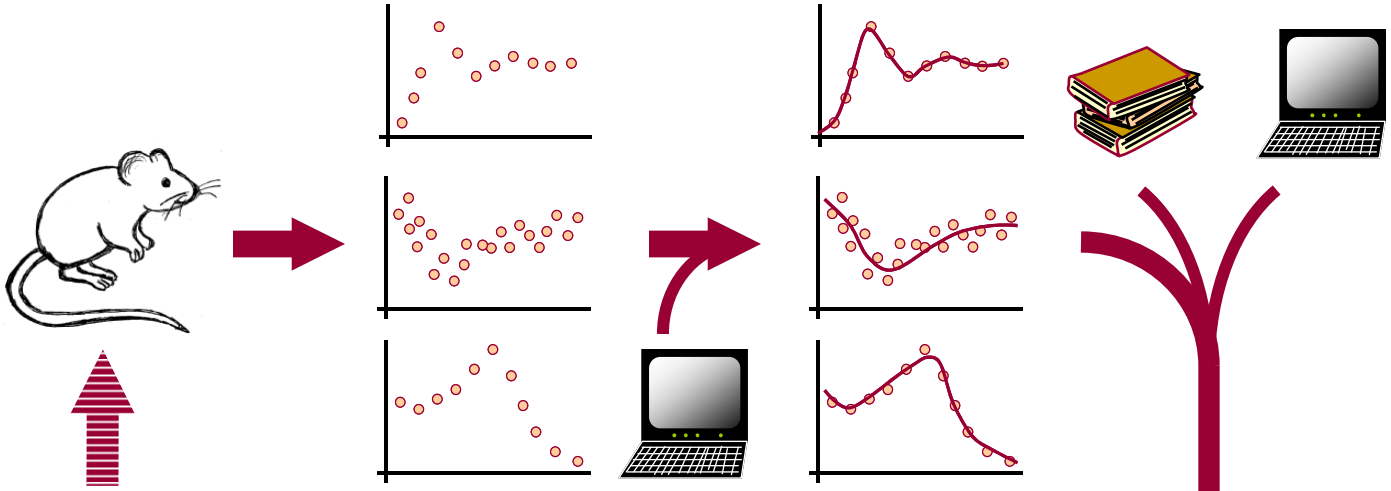

*Voit, Drug Discovery Today*, 2004

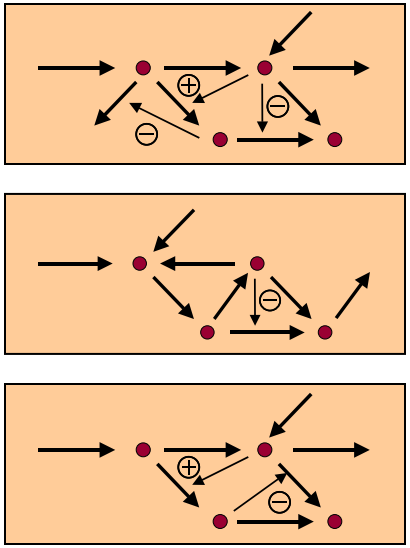
# Alternative to Traditional Modeling: Top-Down Modeling

- Use information at the “global” level (*in vivo* time series data) to deduce (per model) structure and regulation at the “local” level (connectivity, signals,…)



# Top-Down "Inverse" Modeling





$$\dot{X} = \alpha \prod X^{g} - \beta \prod X^{h}$$

$$\dot{Y} = \alpha' \prod Y^{g'} - \beta' \prod Y^{h'}$$

$$\dot{Z} = \alpha'' \prod Z^{g''} - \beta'' \prod Z^{h''}$$

***BST***

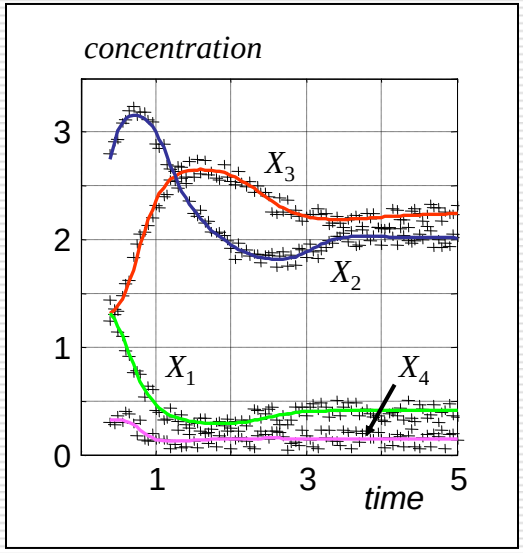
# Key Step: Parameter Estimation from Time Series Data

- According to computer scientists: trivial, solved.
- Many methods
- Most work sometimes
- None works always
- Estimation remains to be a challenging topic!
- Example: Kikuchi *et al.* 2003
- For canonical models, parameter estimation leads to pathway structure identification

# Recent Approaches to Parameter Estimation from Time Series Data

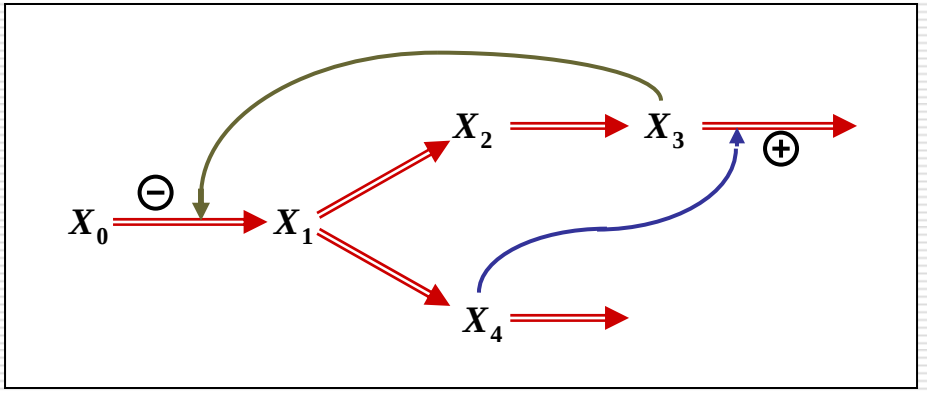
- Substitution of slopes for differentials; including decoupling of equations (Voit and Savageau, 1982)
- Neural networks, Genetic algorithms (U.S.A., Japan, Portugal)
- Nonlinear regression (U.S.A., Spain)
- Interval methods (Switzerland, Sweden, England)
- Newton flow methods (Switzerland, Sweden, England)
- Simulated Annealing (Germany)
- Swarm methods (Philippines)
- Global branch-and-bound methods (U.S.A.)
- Collocation and hybrid evolution (Taiwan)
- Alternating regression (U.S.A., Norway)
- Eigenvector optimization (U.S.A., Brazil)

# Challenge: Parameter Estimation; Structure Identification




VBoMT

Voit's Box of Magic Tricks



# Major Challenge: Ill-defined Pathway Systems

Typical situations:

Not all metabolites and enzymes are known

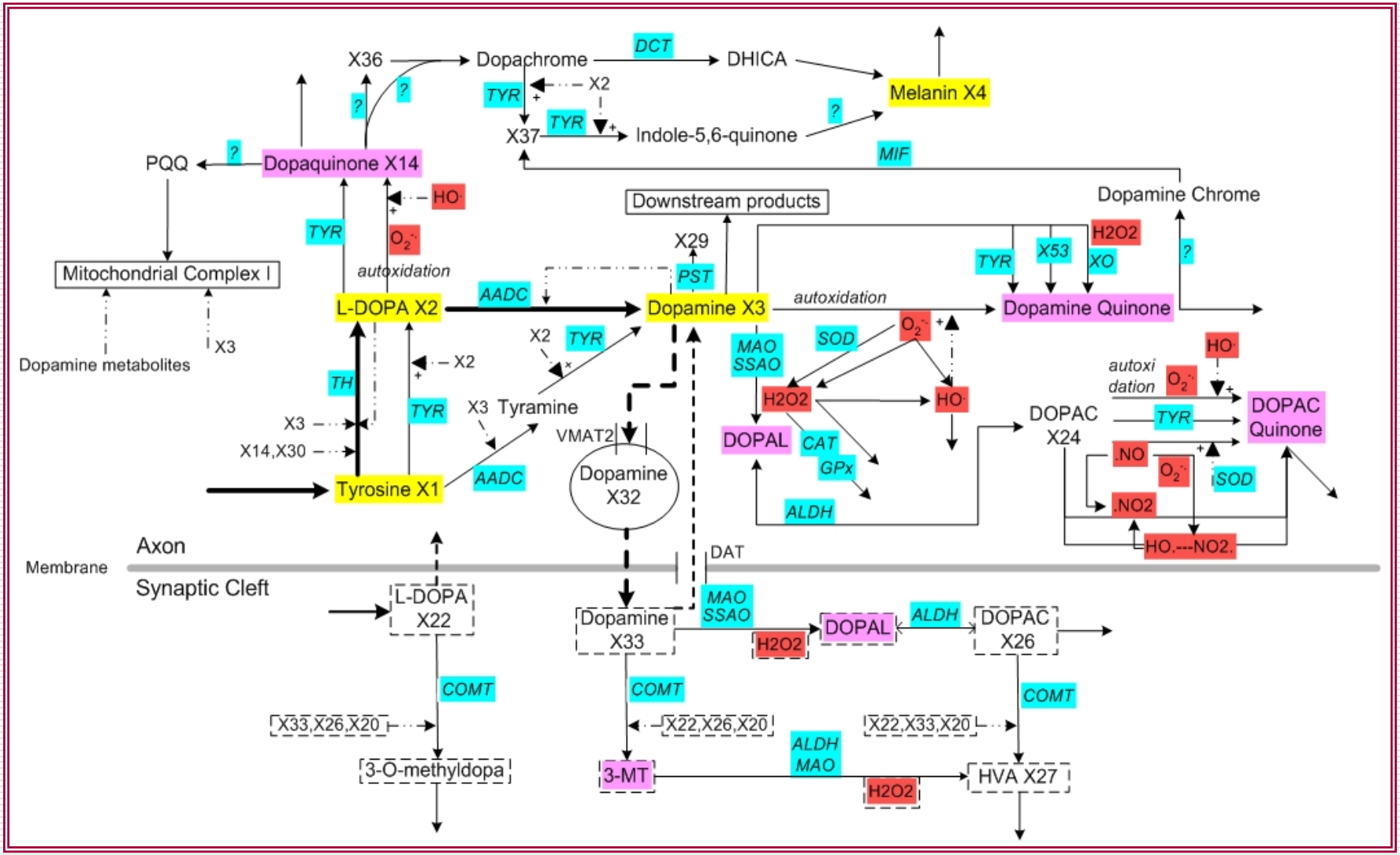
Not all regulatory signals are known

Parameters are uncertain

Concentrations are uncertain

Flux rates are uncertain

# Example: Dopamine Metabolism



# Addressing the Parameter Challenge:

**Expert Opinion**

Ask question in a new fashion:

***Not:*** What is the concentration of DOPAC?
What is the turnover rate of this step?

***But:*** Are concentrations of X and Y about the same?
How do fluxes relate at this branch point?

**Defaults**

Kinetic orders in BST are within small ranges

Restrict ranges based on experience
with enzyme kinetics

**Diagnostics**

Massive sensitivity and robustness analysis

# Illustration: Parkinson's Disease

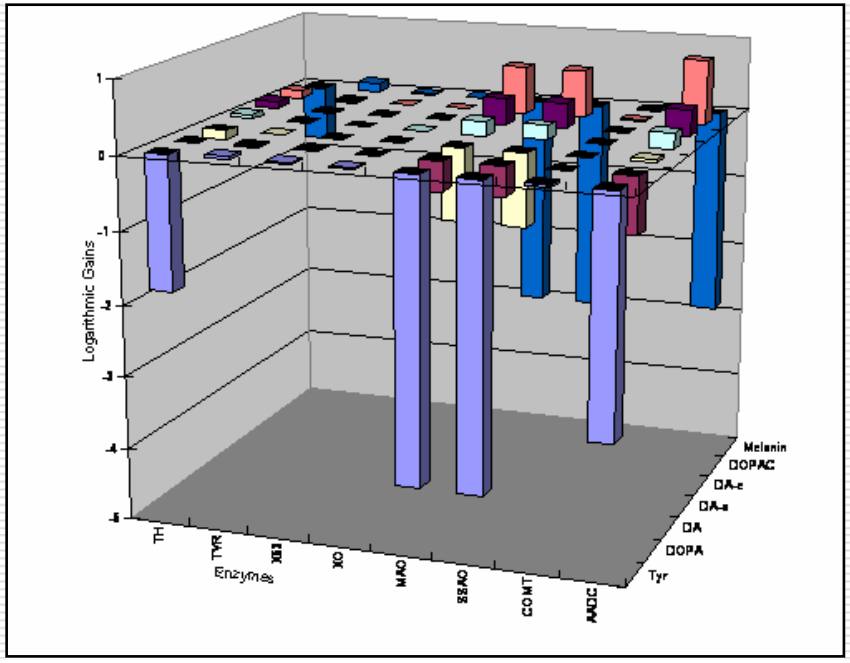
- Collect information on dopamine metabolism
- Discuss topology of the pathway: "network"
- Discuss regulation of the network: "system"
- Construct symbolic systems equations, using Biochemical Systems Theory (BST)
- Guestimate order-of-magnitude concentrations
- Guestimate order-of-magnitude fluxes
- Construct numerical equations
- Diagnose, refine model

Work with Zhen Qi, Gary Miller, Mahlon Delong, and others at Emory University.
Supported by a Predictive Health / Woodruff Foundation Pilot Grant

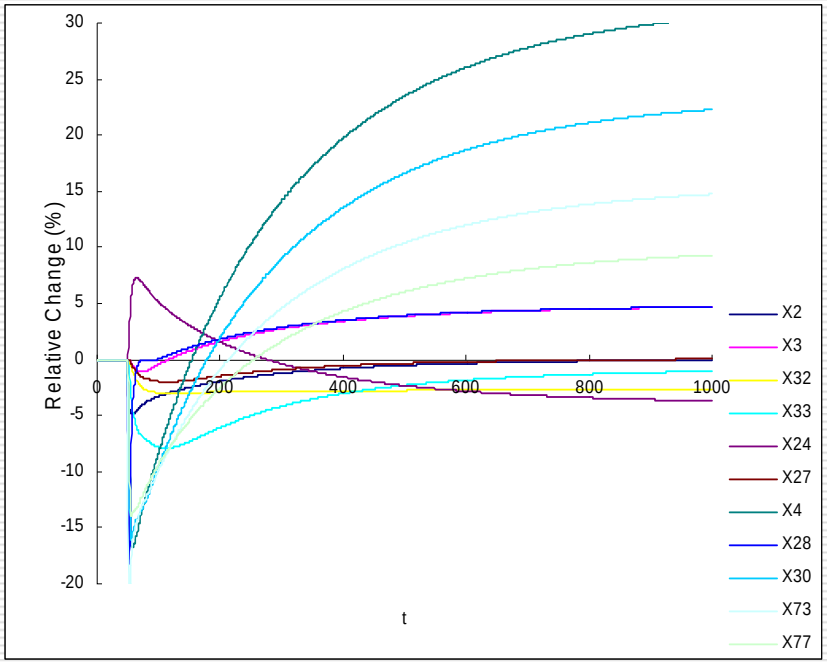
# BST Formulation (excerpt)

```
beta1_01 = a1_0 *J1 /(Xs1^h1_1_01  Xs10^h1_10_01  Xs50^h1_50_01  Xs2^h1_2_01 >>
>>                  Xs3^h1_3_01  Xs14^h1_14_01  Xs30^h1_30_01)
beta1_02 = a1_0 *J2 /(Xs1^h1_1_02  Xs52^h1_52_02  Xs2^h1_2_02)
beta1_03 = a1_0 *J3 /(Xs1^h1_1_03  Xs52^h1_52_03  Xs2^h1_2_03)   //to Dopaquinone # X52
beta1_04 = a1_0 *J4 /(Xs1^h1_1_04  Xs59^h1_59_04  Xs3^h1_3_04)   //to Tyramine # X59
X1' = a1_0    - beta1_01  X1^h1_1_01  X10^h1_10_01  X50^h1_50_01  X2^h1_2_01
>>                 X3^h1_3_01  X14^h1_14_01  X30^h1_30_01 >>
>>                 - beta1_02  X1^h1_1_02  X52^h1_52_02  X2^h1_2_02 >>
>>                 - beta1_03  X1^h1_1_03  X52^h1_52_03  X2^h1_2_03 >>
>>                 - beta1_04  X1^h1_1_04  X59^h1_59_04  X3^h1_3_04
//=============================================================================
alpha10 = beta1_01
beta10 = beta1_01
//X10' = alpha10  X1^h1_1_01  X10^h1_10_01  X50^h1_50_01  X2^h1_2_01  X3^h1_3_01
>>                 X14^h1_14_01  X30^h1_30_01 >>
>>                 - beta10  X1^h1_1_01  X10^h1_10_01  X50^h1_50_01  X2^h1_2_01
>>                 X3^h1_3_01  X14^h1_14_01  X30^h1_30_01
//=============================================================================
h11_11 = 0.5
alpha11 = beta10
beta11 = (alpha11  Xs1^h1_1_01  Xs10^h1_10_01  Xs50^h1_50_01  Xs2^h1_2_01
>>                 Xs3^h1_3_01  Xs14^h1_14_01  Xs30^h1_30_01) /(Xs11^h11_11)
X11' = alpha11  X1^h1_1_01  X10^h1_10_01  X50^h1_50_01  X2^h1_2_01  X3^h1_3_01
>>                 X14^h1_14_01  X30^h1_30_01 >>
>>                 - beta11 X11^h11_11
//=============================================================================
```

# Selected Results




Effects of changes in enzyme activities on metabolite concentrations




Dynamic responses of metabolite concentrations to a 30% decrease in tyrosine hydroxylase

# Selected Results

| Manipulation | Metabolites | Experimental Result | Prediction |
|---|---|---|---|
| **TH heterozygote** | dopamine | No change | -2.68% |
| | DOPAC | No change | 0.78% |
| | HVA | No change | -0.94% |
| **TH knockout** | dopamine | -99.58% | -100.00% |
| | DOPAC | Not detected | -100.00% |
| | HVA | Not detected | -100.00% |
| **COMT heterozygote** | dopamine | 6.93% | 18.56% |
| | DOPAC | 10.54% | 18.93% |
| | HVA | -14.52% | -49.10% |
| **COMT knockout** | dopamine | 10.64% | 37.39% |
| | DOPAC | 232.95% | 464.06% |
| | HVA | -100.00% | -100.00% |
| **COMT heterozygote + 90% DAT inhibition** | dopamine | 21.97% | 14.57% |
| | DOPAC | 71.46% | 89.78% |
| | HVA | -17.01% | -43.36% |
| **COMT knockout + 90% DAT inhibition** | dopamine | 30.58% | 30.25% |
| | DOPAC | 447.50% | 876.77% |
| | HVA | -100.00% | -100.00% |
| **VMAT2 LO[#]** | dopamine | -85.42% | -89.98% |
| | DOPAC | -58.00% | -28.96% |
| | HVA | -58.17% | -83.55% |

# Toward Personalization

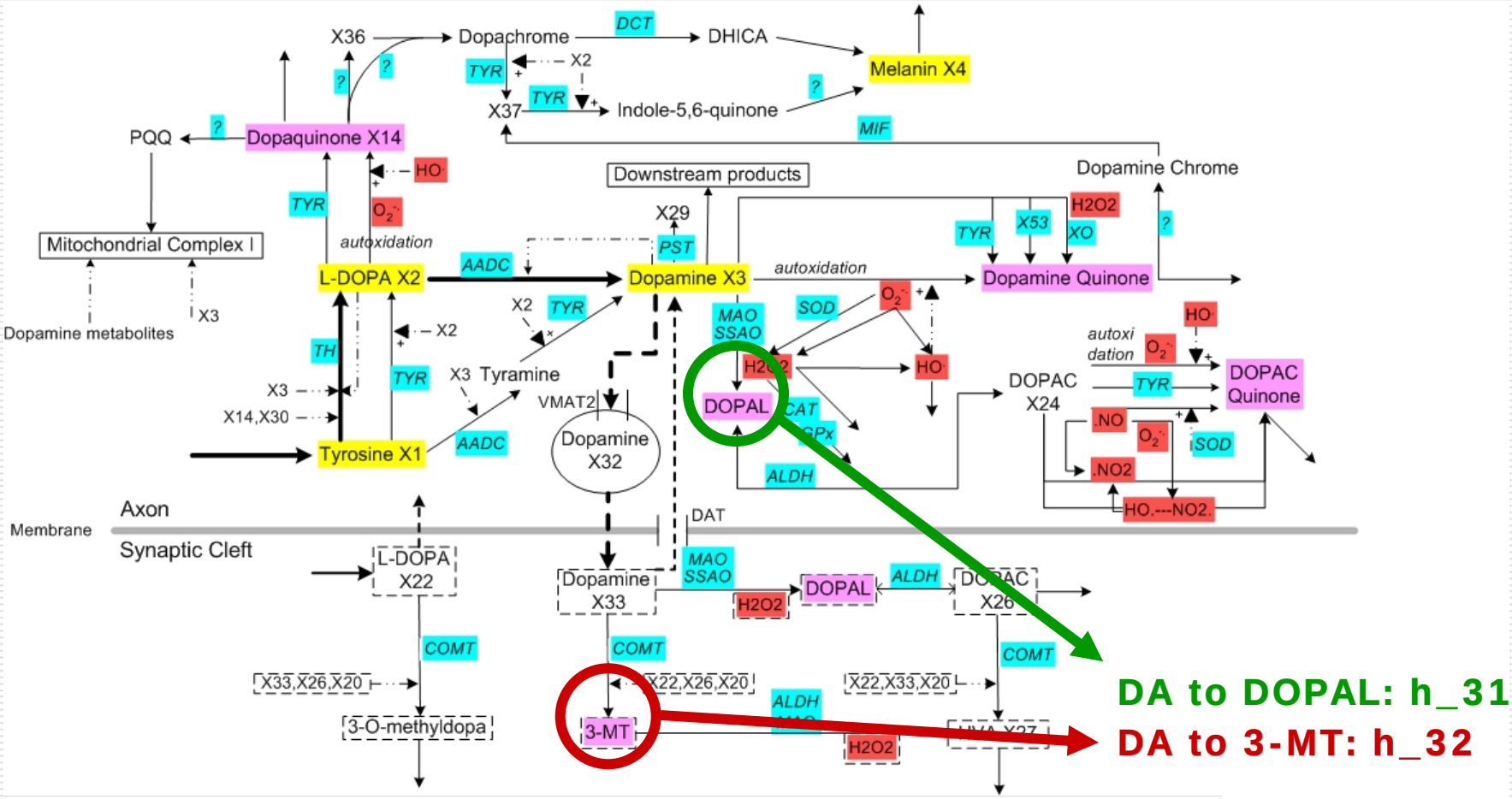
- Pathway diagram shows generic “topology” of dopamine metabolism (including regulation).
- Parameter values (so far) represent average values (*e.g.*, of enzyme activities) in healthy population.
- In reality, enzyme activities and metabolite concentrations exhibit interpersonal variability.
- Assumption is made that the model structure is robust toward this variability (valid for slightly changed parameter values).
- If so, model can be “personalized” by letting some values deviate from population averages.

# Toward Personalization (cont'd)

- Assume for simplicity that enzyme activities vary independently from one another (assumption easy to relax).
- Establish distribution for each enzyme activity and each concentration within a population (measure or math).
- Then: population variability can be assessed by repeatedly drawing a value from each distribution and analyzing model responses ("Monte Carlo simulation").
- Determine "risk factors" through sensitivity analysis.
- Show that some people get sick without exhibiting risk factors and others do not get sick in spite of them.

# Personal Risk Profiles

Illustration (under the assumption that our model is correct):
Study increases in toxic species 3-MT and dopamine quinone)



# Personal Risk Profiles

| Person | h3_1 | h3_2 | 3-MT | Dopamine Quinone |
|---|---|---|---|---|
| 1 | √ | √ | √ | √ |
| 2 | ⇓ 20% | √ | ⇓ 13% | ⇑ 70% |
| 3 | √ | ⇑ 20% | ⇑ 120% | ⇓ 48% |
| | **Question: Combined Effect?** | | | |
| 4 | ⇓ 20% | ⇑ 20% | ⇑ 127% | ⇓ 31% |

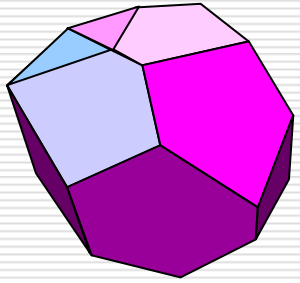
# Computation of All Risk Profiles

Attempt to look at singular and multiple “aberrations” and studied output. ➡ “Combinatorial Explosion”

BST modeling allows elegant assessment of ***all healthy*** and ***all pathological profiles***.

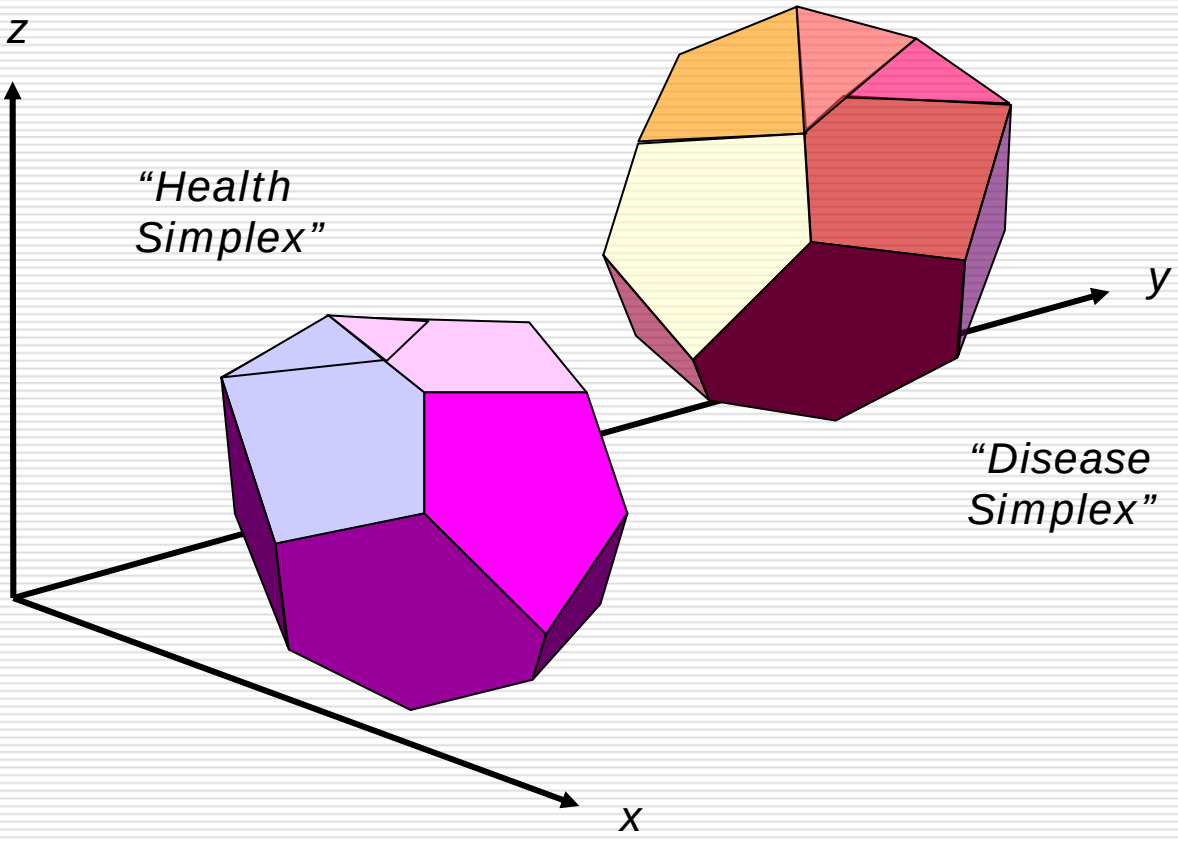
Mathematical approach: Using BST (S-system) representation, invert steady-state matrix or perform straightforward (linear) constrained optimization.

Solution: One “Simplex” each.



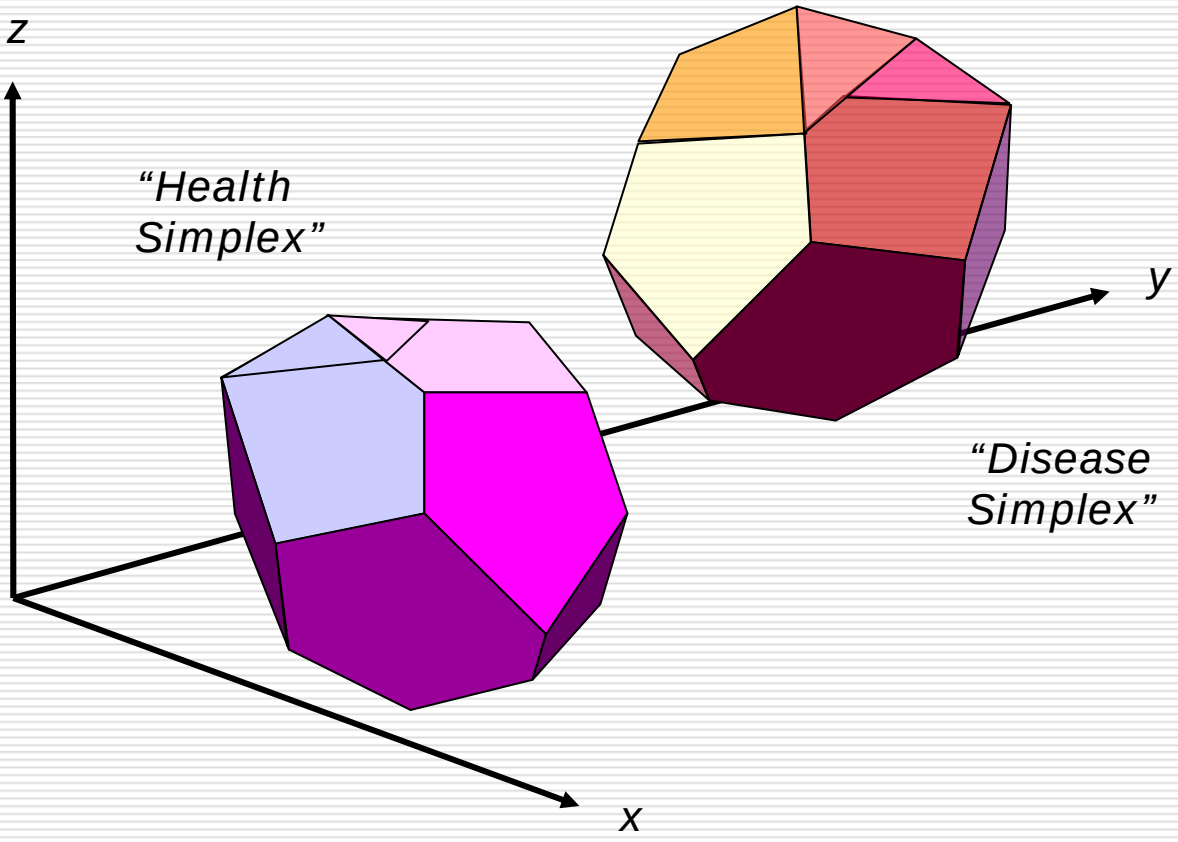
# Computation of All Risk Profiles

**Ideal Solution (in full "biomarker space"):**
**Clear separation between health and disease simplexes**



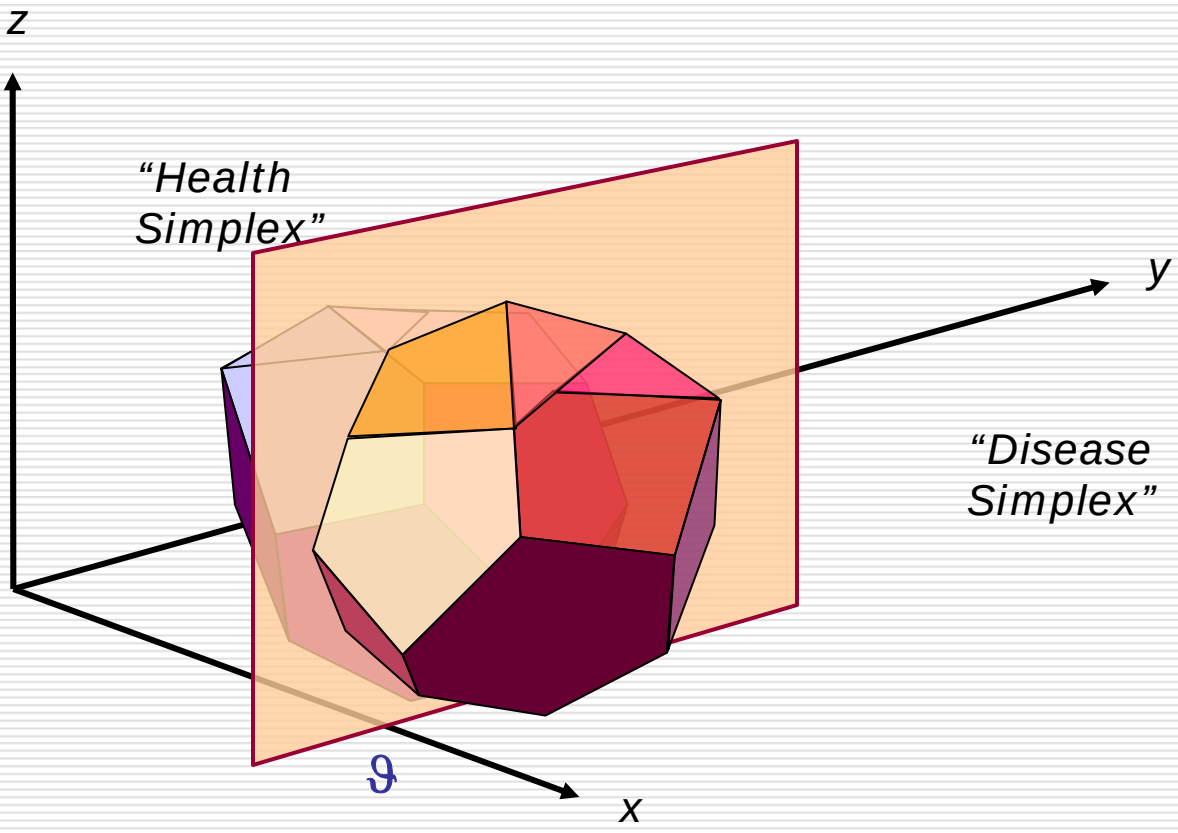
# Computation of All Risk Profiles

## Realistic Situation:
## Projection onto a lower-dimensional biomarker space



# Computation of All Risk Profiles

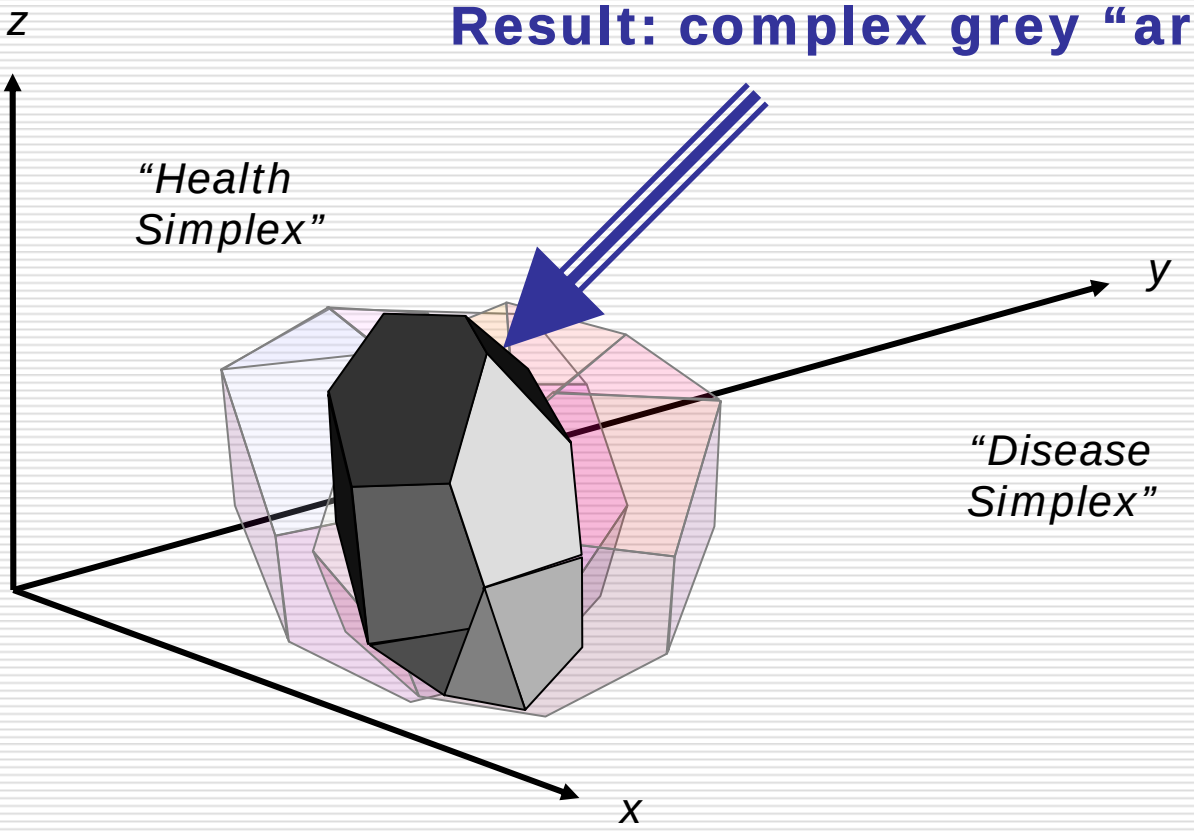
**Would like to say: $x < \vartheta$ : healthy; $x > \vartheta$ : sick (like PSA > 4)**



# Computation of All Risk Profiles

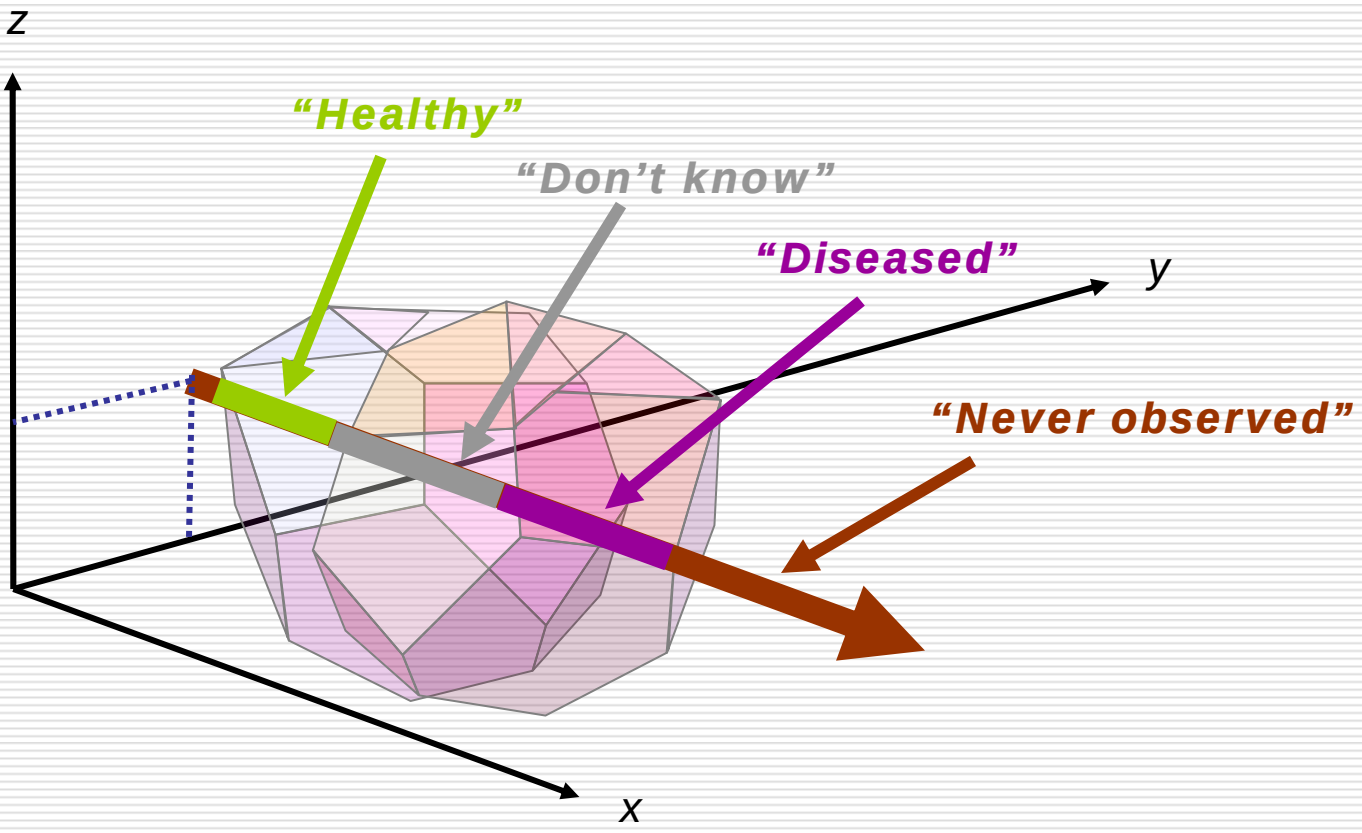
**In reality, *there is no unique* $\vartheta$ because disease status also depends on other biomarkers, such as y and z.**

**Result: complex grey "area"**



# Computation of All Risk Profiles

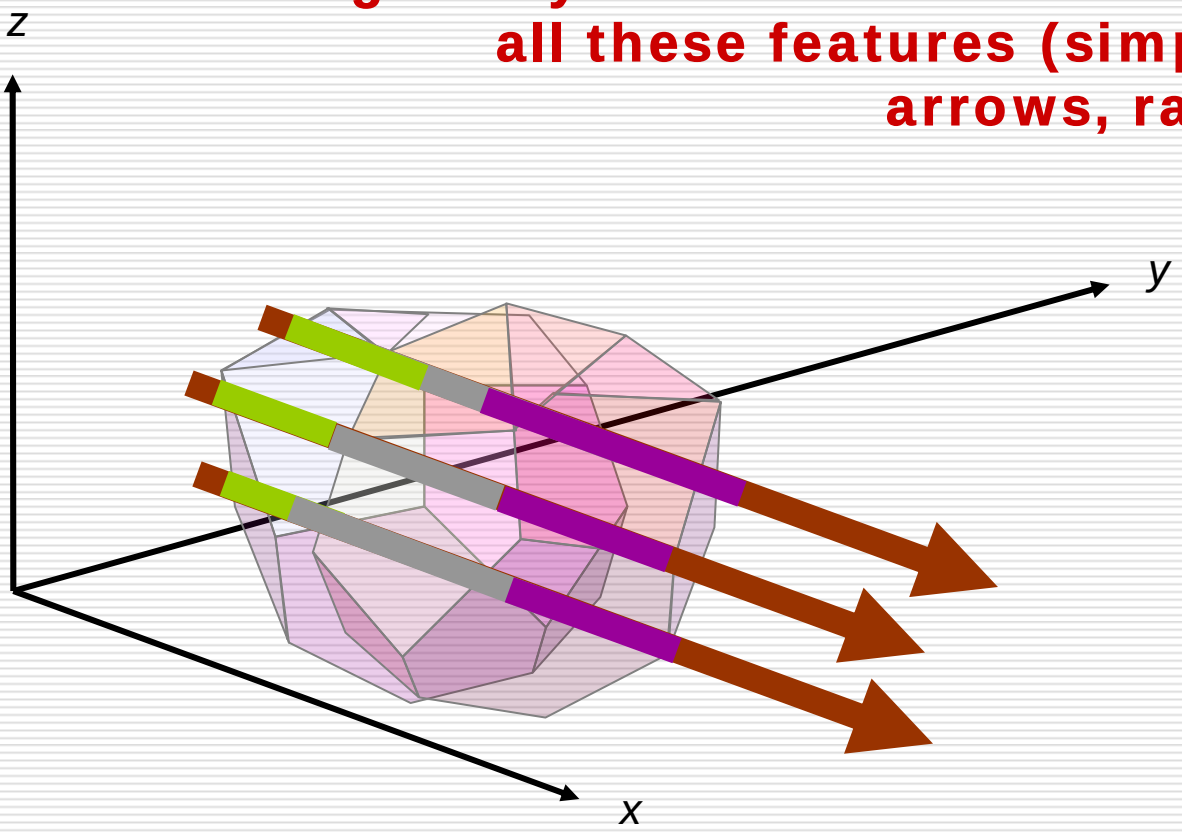
**Consequence: Looking at one biomarker insufficient.**



# Computation of All Risk Profiles

**"Don't know" range depends on (unobserved) y and z.**

**A good systems model characterizes all these features (simplexes, arrows, ranges).**



# Opportunities

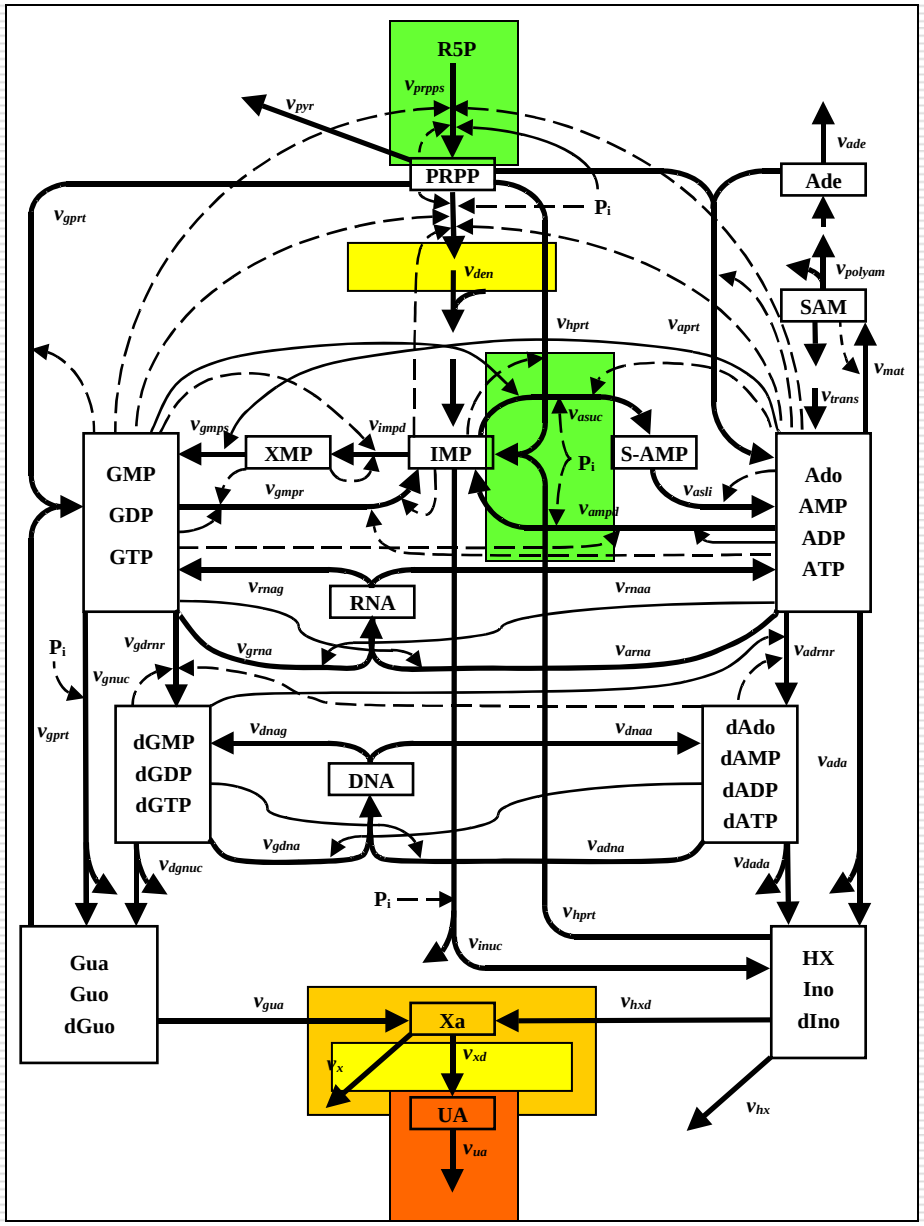
Extend from metabolism to physiology

Define “health” and “disease”

Define “biomarker” and “pre-morbid”

Characterize trajectories toward disease

Study personalized treatment regimens




# Personalized Treatment (Hyperuricemia)

Suppose too much UA

1. Explain:
   e.g., PRPPS superactivity
   or, HGPRT deficiency
2. Intervene:
   Allopurinol
   6-Thioguanine
3. Side effects?
   e.g.: UA ⇓ ⇒ Xa ⇑

# Degrees of Difficulty for Systems Biology Based Medicine

| *Activity* | *Difficulty* |
| --- | --- |
| Model equations |  |
| Parameter values |  |
| Model analysis |  |
| Simulation |  |
| Prediction |  |
| Optimization |  |
| Implementation |  |

# Summary

For Leslie Wesley it may be too late.

But hopefully, in two thousand and twenty eight,

We have enough data, computers and wisdom

To figure out this complex system.

# Acknowledgments

**Zhen Qi**

**Gary Miller & Mahlon Delong**

**Emory DISCOVER Project Team**

**Woodruff Foundation**

# Acknowledgements

***The Current Crew:***

***Funding: NIH, NSF, DOE, Woodruff Foundation, Georgia Research Alliance***

***Information: www.bst.bme.gatech.edu***